Plant natural products: the molecular genetic basis of biosynthetic diversity
Richard A Dixon

Significant advances have been made concerning the biosynthesis, regulation and genetic manipulation of plant natural products. These include insights into the structural biology of isoprenoid cyclization and polyketide condensation reactions, a better understanding of the molecular biology of plant cytochrome P450s and O-methyltransferases, and new information on the effects of natural products on human health.

Introduction
The vast array of plant natural products can be viewed as the current status of nature’s activities in combinatorial biochemistry. Understanding the molecular biology of their often complex biosynthetic pathways will lead to new opportunities for improving plant disease resistance, increasing levels of human health promoting compounds (nutraceuticals) in food crops, and utilizing transgenic plants for production of pharmaceuticals [1].

Much of the research on plant natural product biosynthetic pathways is still at the level of gene discovery. The recent studies reviewed here have increased our understanding of how biosynthetic diversity is generated using relatively conserved enzymatic mechanisms. This emerging knowledge will provide the bridge to the next era of natural product gene discovery utilizing functional genomics.

Isoprenoid biosynthesis
Following the discovery that eubacteria can produce isoprenoids (natural products derived from acetate via the C₅ isoprene unit) by a pathway that does not involve 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR) [2], it became increasingly apparent that plants might have more than one pathway to isopentenyl diphosphate (IPP), the immediate precursor of the terpenoids. For example, the patterns of incorporation of ¹³C-glucose and ¹³C-acetate into taxanes in Chinese yew cell cultures suggest that the anti-cancer drug taxol is synthesized by a bacterial-type isoprenoid pathway, in which IPP originates from 1-deoxyxylulose 5-phosphate (DXP). DXP is itself made from pyruvate and glyceraldehyde 3-phosphate (GAP) by a specific transketolase enzyme [3]. These discoveries have led to a major re-evaluation of the early stages of isoprenoid biosynthesis. Many, if not all, plastid-derived plant isoprenoids, including monoterpenes, diterpenes, isoprene, carotenoids and the prenyl side chains of chlorophyll and plastoquinone, are now believed to originate from pyruvate and GAP [4••]. The transketolase (DXP synthase) has been cloned from peppermint by mass sequencing of a secretory gland cDNA library [5••], from pepper by PCR based on Arabidopsis expressed sequence tags (ESTs) [6•], and from Escherichia coli, where the gene is in an operon linked to the gene encoding farnesyl diphosphate synthase, another enzyme involved in isoprenoid biosynthesis [7]. DXP synthase is localized in plastids in plants, and is particularly highly expressed in chromoplasts that make isoprenoid pigments [6•].

Previous studies on the genetic manipulation of HMGR in plant cells reported effects on sterol metabolism, for which HMGR appears to be a key rate-limiting enzyme [8,9]. The existence of the DXP pathway might explain why changes in other isoprenoid-derived products were not observed in HMGR transgensics, although potential metabolic cross-talk between the two pathways may occur [4••].

Specific terpene cyclases define the folding patterns of the allylic terpenoid precursors. There is now sufficient information on conserved sequence motifs between mono-, sesqui- and di-terpene cyclases [10••] to facilitate cloning by PCR, as recently demonstrated for the germacrene C synthase from tomato [11] and the (E)-β-bisabolene synthase involved in the formation of insect juvenile hormone mimics from grand fir [12]. Furthermore, the ability to mass sequence cDNA libraries from tissues, such as oil glands, that are dedicated to the synthesis of terpenoids has made it possible, knowing signature sequences, to rapidly select candidate clones for functional expression. In this way, the (E)-β-farnesane synthase that converts farnesyl diphosphate to the aphid alarm pheromone was cloned from peppermint [13•]. The increasing amount of primary sequence information for functionally characterized terpene synthases, coupled with the recent publication of the crystal structures of tobacco 5-epi-aristolochene synthase (a sesquiterpene cyclase) [14••], now open up the possibility of rational design of novel cyclase enzymes by site-directed mutagenesis or domain swapping approaches to introduce novel regiochemical and stereochemical diversity and thus generate new bioactive natural products in transgenic plants.
Polyketide synthases

Chalcone synthase (CHS), the first plant natural product polyketide synthase to be characterized at the molecular level [15], catalyzes the head-to-tail condensation of 4-coumaroyl CoA with three molecules of malonyl CoA to yield naringenin chalcone, a precursor of the major classes of plant flavonoids. An alternative folding pathway for the final enzyme-associated polyketide product leads instead to the formation of the stilbene resveratrol, and CHSs and stilbene synthases are so closely related at the primary sequence level that they do not fall into distinct families [16]. It is now apparent that CHS is a member of a family of closely related polyketide synthases that can utilize different starter molecules, and perform different numbers of condensation reactions, to yield a variety of interesting and potentially valuable natural products, such as benzalacetones, acridones, styrylpyrones, benzophenones leading to xanthones, and benzenes, as well as chalcones and stilbenes with different substitution patterns [17••,18,19] (see Figure 1). It is probable that a significant number of the enzymes appearing in the databases as CHS in fact encode other related polyketide synthases.

This emerging picture of catalytic flexibility among the CHS-related plant polyketide synthases recently received further support from the demonstration that the product of a gene, originally called gch2, from the ornamental Gerbera hybrida (Asteraceae), catalyzed the formation of 6-methyl-4-hydroxy-2-pyrene from one molecule of acetyl CoA (the primer) and two molecules of malonyl CoA [20••] (see Figure 1). This new function was revealed by the observation that Gerbera plants expressing antisense gCHS2, which has 73% amino acid sequence identity to functionally confirmed Gerbera CHS, completely lack two prominent pyrone derivatives, gerberin and parasorboside.

Understanding the molecular basis of the starter specificity, condensation reactions, chain termination and polyketide cyclizations of CHS-like enzymes may facilitate production of designer enzymes for formation of novel plant polymers. The complete crystal structure of an alfalfa CHS has recently been solved at the 1.58–1.8 Å level (JL Ferrer et al., unpublished data), and this now provides a rational basis for testing hypotheses on reaction modification by site-directed mutagenesis.

Transgenic expression of natural or genetically modified CHS-family enzymes in plants will find applications in engineering defense compounds and nutraceuticals. For example, the stilbene resveratrol, long known to be a phytoalexin active against a range of plant pathogens, is now recognized as having significant anticancer activity in animal mammary and skin cancer models [21••], possibly as a result of reduced DNA synthesis due to its inhibitory activity against ribonucleotide reductase [22]. Resveratrol synthesis can be readily engineered into plants by gene transfer technology ([23]; J Hipskind, NL Paiva, personal communication).

Cytochrome P450 enzymes

Cytochrome P450s are heme-containing enzymes that catalyze many, but not all, of the hydroxylation reactions of plant natural product biosynthesis, as well as more ‘exotic’ reactions, such as phenol coupling and oxidative ring closure [24••]. In a recent review [25•], 57 published plant cytochrome P450 sequences were listed, of which 28 encode enzymes of unknown function. This list does not include the large number of cytochrome P450 sequences now appearing in EST databases. Major advances in this area have come from the employment of PCR-based strategies for generation of candidate cytochrome P450 sequences, and the use of heterologous systems, such as yeast, E. coli, and insect cells, for functional expression of cloned plant cytochrome P450s.

Although primary sequences of cytochrome P450 enzymes are highly divergent, conserved motifs, including the heme binding region and a proline-rich motif downstream of the amino-terminal membrane anchor, provide sufficient homology for design of degenerate primers for PCR amplification of multiple cytochrome P450 DNA sequences from a single tissue. In this way, 16 different, but as yet functionally uncharacterized, cytochrome P450s were amplified from an Arabidopsis thaliana cDNA library, each demonstrating unique tissue specificity and responsiveness to wounding and light/dark treatments [26•]. A similar strategy, coupled to heterologous expression, has resulted in the identification of CYP71E1, which catalyzes the two-step conversion of p-hydroxyphenyl-acetaldoxime to p-hydroxymandelonitrile in the biosynthesis of the anti-herbivore cyanoecigenic glycoside dhurrin in Sorghum bicolor [27••]. The first six reactions in the conversion of L-tyrosine to dhurrin are catalyzed by just two cytochrome P450s [28•], a good example of metabolic channeling in natural product biosynthesis.

Other interesting plant cytochrome P450s have recently been functionally characterized in either yeast and/or insect cells: CYP93B1, the (2S)-flavanone 2-hydroxylase involved in the flavanone to flavone or retrochalcone conversion in licorice (Glycyrrhiza echinata L.) [29]; CYP81E1, the isoflavone 2'-hydroxylase from the same source [30]; CYP76B1, a xenobiotic-inducible 7-ethoxycoumarin-O-de-ethylase from sunflower (Helianthus tuberosus) that may be useful as a marker for bio-monitoring of pollution [31]; and CYP80B1, a methyl jasmonate-inducible enzyme catalyzing the 3'-hydroxylation of (s)-N-methylcoclaurine in the biosynthesis of analgesic and anti-microbial benzylisoquinoline alkaloids in California poppy (Eschscholzia californica) [32]. Isoflavone 2'-hydroxylase may be the rate limiting step in the formation of antimicrobial pterocarpan phytoalexins in chickpea [33], suggesting that over-expression of this enzyme might be a potential strategy for improving disease resistance in legumes.

2,4-Dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA) is a cyclic hydroxamic acid that confers on corn resistance to both insects and fungal pathogens. A combination of
transposon tagging with the mutator element, differential library screening and genetic mapping techniques has revealed that DIMBOA is synthesized from indole by four consecutive cytochrome P450 reactions [34••]. The first hydroxylation is catalyzed by a cytochrome P450 that is located within 2.5 kb of the tryptophan synthase α-homolog that forms the precursor indole, the final N-hydroxylase in the pathway is located within 4 cM of the genes encoding these two enzymes, and the two central cytochrome P450s in the pathway are linked 2 cM from the N-hydroxylase. It will be interesting to map the loci of the many cytochrome P450 genes now represented as ESTs in the Arabidopsis database in order to obtain a general picture of the extent of clustering and, therefore, potential evolution of the plant cytochrome P450 superfamily by gene duplication.

The electron donor for cytochrome P450 reactions is NADPH, with electrons being passed to the cytochrome P450 heme group via NADPH cytochrome P450 reductase. Although endogenous cytochrome P450 reductase(s)
can function during the heterologous expression of plant cytochrome P450s, optimum activity may depend on reconstitution with the homologous reductase [35*].

**Phenylpropanoid biosynthesis**

In addition to lignin monomers and lignans, many of the flavonoids, isoflavonoids and other hydroxycinnamic-acid-derived natural products contain methoxyl substituents, formed by the action of highly regiospecific O-methyltransferases (OMTs). Many of these OMTs have been cloned following classical biochemical purification, and over 50 plant OMT sequences now exist in the databases, allowing analysis of evolution and structure-function relations [36*]. Unexpectedly, unlike the lignin OMTs of most plants, two OMTs from the saxifrage *Chryosplenium americanum* catalyze the 3′-O-methylation of the flavonoids luteolin and quercetin as well as the 3/5-O-methylation of the lignin precursors caffeic and 5-hydroxyferulic acids [37]. In *Clarkia breweri*, the isoeugenol OMT involved in floral scent production has recently been cloned and domain swapping experiments have revealed the sequence elements that differentiate the substrate specificities of lignin-type caffeic acid OMT and isoeugenol OMT [38*].

The antifungal isoflavonoids of alfalfa and chickpea contain a B-ring methoxyl group, although it has not been possible to date to isolate an enzyme that can catalyze its formation *in vitro*. However, an OMT that methylates the A-ring hydroxyl of the isoflavone precursor daidzein, to yield isoflorononetin, has been cloned and proposed to catalyze the B-ring methylation reaction *in vivo* [39*].

Interest in isoflavonoids as nutraceuticals has been further strengthened by the reports that the soybean isoflavones genistin and daidzein, in addition to exhibiting estrogenic effects on improving bone mass [42]. Of the enzymes necessary for engineering isoflavone nutraceuticals into plants, only two, the 2-hydroxylase and dehydratase of the isoflavone synthase complex, have yet to be characterized at the molecular level. The dimeric lignans similarly have potent anticancer and antioxidant activity [39*], and genes encoding all the enzymes for the conversion of coniferyl alcohol to secoisolariciresinol, a major dietary phytoestrogen, have been cloned. These include the remarkable dirigent protein that co-acts with oxidases to confer stereochemical free radical coupling [43**], and the (±)-pinoresinol/(±)-lari-
ciresinol reductase that shares extensive sequence similarity to legume isoflavone reductases [44].

Utilizing CHS and dihydroflavonol reductase constructs, it has been possible to alter the content and composition of condensed tannins in birdsfoot trefoil (*Lotus corniculatus*) [45–47]. These studies are important because condensed tannins are believed to help prevent bloat in ruminants feeding on highly digestible forages. More global upregulation of phenylpropanoid biosynthesis by over-expression of L-phenylalanine ammonia-lyase results in increased local and systemic resistance of tobacco to microbial pathogens, but compromised systemic resistance to herbivorous insect larvae [48]. This underlines the potential for unexpected metabolic cross-talk during genetic manipulation of natural product pathways.

**Conclusions**

The amount of molecular information on several classes of plant natural product biosynthetic enzymes, such as terpene cyclases, polyketide synthases, cytochrome P450s and O-methyltransferases, is now sufficient to facilitate cloning the corresponding genes from any plant species by PCR strategies, or by EST database searching. Functional analysis in heterologous systems, such as *E. coli*, yeast or insect cells, is becoming routine, and so genomics-based approaches can now complement, or in many cases replace, classical enzyme purification strategies for isolation of commercially important natural-product pathway genes. The challenge now is to develop methods for *in planta* functional analysis in those species that are rich sources of natural products but are genetically intractable.

**Acknowledgments**

Work in the author’s laboratory was supported by the Samuel Roberts Noble Foundation.

**References and recommended reading**

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest


An excellent review of the history of the discovery of the Rohmer pathway for synthesis of isoprenoids, the reaction mechanisms of the pathway itself, the involvement of deoxyxylulose in the synthesis of non-terpenoid natural products, and the evolution of the pathway. Includes a list of plant and bacterial natural products whose synthesis can unequivocally be assigned to the Rohmer pathway or the mevalonate pathway.


The authors describe the cloning from peppermint of cDNAs encoding the key enzyme of the Rohmer pathway of isoprenoid biosynthesis. The enzyme has a typical transit peptide for transport to plastids. Clones were identified by mass sequencing of an epidermal oil gland cDNA library, a strategy that also led to the cloning of the β-farnesene synthase (see [13*]).


Two transketolases were identified from pepper, one encoding a plastidial enzyme of the pentose phosphate pathway, the other the DXP synthase of
the Rhomer pathway, the latter also being involved in biosynthesis of thiamine and pyridoxine.


A review of the structural comparisons between mono-, sesqui- and di-terpene synthases, the mechanisms of terpene cyclization, and the evolution of the various terpene synthase families.


12. Bohlmann J, Crock J, Bettcher A, Schröder G, Stöcker R: The authors describe the cytochrome P450s and 2-oxoglutarate-dependent dioxygenases involved in the biosynthesis of trihydrobenzisoxazoline, and monoterpenoid indole and tropane alkaloids. These enzymes often catalyze complex reactions, such as phenol coupling, that are difficult to mimic chemically with retention of regio- or stereo-specificity. These enzymes are major targets for biotechnological exploitation.


The crystal structure of a tobacco sesquiterpene cyclase was determined at 2.2 Å resolution, and in complexes with farnesyl diphosphate derivatives. This provides the basis for modeling the reaction mechanisms of all terpene cyclases, and for the design of cyclases with novel reaction pathways.


An excellent review of the biochemical flexibility of the family of plant polyketide synthases related to chalcone synthase. Highlights the dangers of ascribing function to cloned enzymes of this family based on sequence information alone.


20. Eckermann S, Schroeder G, Schmidt J, Strack D, Erdra RA, Helarutta Y, Eimala P, Koltainen M, Kipitainen J, Proksch P et al.: New pathway to polyketides in plants. *Nature* 1998, 396:387-390. The authors report a new member of the chalcone synthase-like gene family that uses acetyl-CoA as primer in the synthesis of pyrones. It can also utilize benzoyl-CoA as primer to synthesize the backbone of phytopyrones, synthetic derivatives of which are inhibitors of HIV-1 protease activity. This work indicates that chalcone synthase-related enzymes are involved in the synthesis of a much larger range of natural products than at first expected.


This paper describes the activity of the phytoalexin resveratrol as an antioxidant, antimutagen, and anti-inflammation, antiproliferation and antipropulsion agent in animal. Cancer model systems. Resveratrol has been suggested to be one of the major health-promoting agents in red wine, and is a natural product that can be readily introduced into plants by genetic engineering.


The authors describe the cytochrome P450s and 2-oxoglutarate-dependent dioxygenases involved in the biosynthesis of trihydrobenzisoxazoline, and monoterpenoid indole and tropane alkaloids. These enzymes often catalyze complex reactions, such as phenol coupling, that are difficult to mimic chemically with retention of regio- or stereo-specificity. These enzymes are major targets for biotechnological exploitation.


An overview of the status of the molecular analysis of plant cytochrome P450s, with detailed discussions of selected examples from the phenylpropanoid, jasmonate, steroid, alkaloid and gibberellin pathways.


An example of the power of the PCR approach for the cloning of novel cytochrome P450s.

27. Bak S, Kahn RA, Nielsen HL, Moller BL, Balk J: Cloning of three A-type cytochromes P450, CYP71E1, CYP98, and CYP99 from *Sorghum bicolor* (L.) meyenboc by a PCR approach and identification by expression in *Escherichia coli* of CYP71E1 as a multifunctional cytochrome P450 in the biosynthesis of the cyanogenic glucoside dhurrin. *Plant Mol Biol* 1998, 36:393-405. A further example of the PCR approach for cloning P450s. Reports the reconstitution in vitro of the whole membrane-associated pathway leading to the cyanogenic glucoside dhurrin utilizing two recombinant P450s, CYP79 and CYP71E1. This is one of the best characterized examples of metabolic channeling in plant secondary metabolism, and is reviewed in more detail in [28*].


An excellent review of the biochemical flexibility of the family of plant polyketide synthases related to chalcone synthase. Highlights the dangers of ascribing function to cloned enzymes of this family based on sequence information alone.


A seminal paper reporting the molecular elucidation of the pathway of DIM-BOA synthesis in corn by a combination of transposon tagging and differential library screening approaches. This demonstrates the involvement of four consecutive cytochrome P450s, and genetic linkage of the enzymes of the pathway.


Co-expression of a cytochrome P450 reductase from opium poppy along with the C-O phenol coupling enzyme of bisbenzylisoquinoline alkaloid biosynthesis (CYP80A1) in insect cells leads to a different product profile from that obtained when relying on the endogenous insect cell reductase in the absence of the plant enzyme.


The authors compare 36 O-methyltransferase sequences, representing genes involved in lignin, flavonoid, furanocoumarin and alkaloid biosynthesis. Defines conserved regions involved in substrate binding, and proposes an evolutionary scheme based in part on substrate preference.


The first report of engineering altered plant OMT substrate- and regio-specificity by domain swapping.


Defines conserved regions involved in substrate binding, and proposes an evolutionary scheme based in part on substrate preference.


This paper reports the cloning of an O-methyltransferase that may exhibit different regiospecificity in vivo and in vitro. Genomics-based identification of such enzymes will therefore require functional confirmation in planta rather than in the usual heterologous systems.


